



Cai, C., Marsh, A., Zhang, Y-H., & Reid, J. P. (2017). Group Contribution Approach To Predict the Refractive Index of Pure Organic Components in Ambient Organic Aerosol. *Environmental Science and Technology*, 51(17), 9683-9690.  
<https://doi.org/10.1021/acs.est.7b01756>

Publisher's PDF, also known as Version of record

License (if available):  
CC BY

Link to published version (if available):  
[10.1021/acs.est.7b01756](https://doi.org/10.1021/acs.est.7b01756)

[Link to publication record in Explore Bristol Research](#)  
PDF-document

This is the final published version of the article (version of record). It first appeared online via ACS at <http://pubs.acs.org/doi/abs/10.1021/acs.est.7b01756>. Please refer to any applicable terms of use of the publisher.

## University of Bristol - Explore Bristol Research

### General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available:  
<http://www.bristol.ac.uk/red/research-policy/pure/user-guides/ebr-terms/>

# Group Contribution Approach To Predict the Refractive Index of Pure Organic Components in Ambient Organic Aerosol

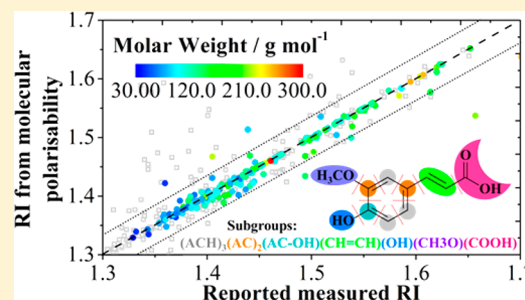
Chen Cai,<sup>†,‡</sup> Aleksandra Marsh,<sup>†</sup> Yun-hong Zhang,<sup>‡,§</sup> and Jonathan P. Reid<sup>\*,†,§</sup>

<sup>†</sup>School of Chemistry, University of Bristol, Bristol BS8 1TS, United Kingdom

<sup>‡</sup>The Institute for Chemical Physics, Beijing Institute of Technology, Beijing 100081, People's Republic of China

<sup>§</sup> Supporting Information

**ABSTRACT:** We introduce and assess a group contribution scheme by which the refractive index (RI) ( $\lambda = 589$  nm) of nonabsorbing components common to secondary organic aerosols can be predicted from the molecular formula and chemical functionality. The group contribution method is based on representative values of ratios of the molecular polarizability and molar volume of different functional groups derived from data for a training set of 234 compounds. The training set consists of 106 nonaromatic compounds common to atmospheric aerosols, 64 aromatic compounds, and 64 compounds containing halogens; a separate group contribution model is provided for each of these three classes of compound. The resulting predictive model reproduces the RIs of compounds in the training set with mean errors of  $\pm 0.58$ ,  $\pm 0.36$ , and  $\pm 0.30\%$  for the nonaromatic, aromatic, and halogen-containing compounds, respectively. We then evaluate predictions from the group contribution model for compounds with no previously reported RI, comparing values with predictions from previous treatments and with measurements from single aerosol particle experiments. We illustrate how such comparisons can be used to further refine the predictive model. We suggest that the accuracy of this model is already sufficient to better constrain the optical properties of organic aerosol of known composition.



## I. INTRODUCTION

Atmospheric aerosols are important in determining the global radiative balance through scattering and absorbing solar radiation.<sup>1–8</sup> Organic material partitioned into the condensed phase<sup>6,9</sup> associated with accumulation and coarse mode primary and secondary aerosol particles effectively scatters and absorbs incoming solar radiation and outgoing terrestrial radiation as well as influences the lifetime and albedo of clouds.<sup>8</sup> The chemical composition of particles determines their complex refractive index (RI), thereby influencing the efficiency of their interaction with light of a given wavelength and the ensuing impacts on climate, altering surface irradiances and actinic flux,<sup>10</sup> and on ozone formation through the reduction in surface UV radiation.<sup>11,12</sup> Commonly, atmospheric aerosols are composed of both inorganic and organic compounds, with compositions that vary greatly in both space and time.<sup>13,14</sup> The RIs of many inorganic aerosol components have been reasonably well established.<sup>15</sup> In contrast, the RIs of organic aerosol components have not been so completely characterized.

Current quantification of the chemical composition of the organic aerosol fraction is limited by its chemical complexity,<sup>10</sup> with hundreds (even thousands) of individual chemical compounds simultaneously present in the aerosol phase compared to the much simpler case of inorganic species. Until recently, the number of direct measurements of the RI of ambient aerosol was also rather limited.<sup>16–21</sup> It is incredibly difficult to infer the compounds present in ambient aerosol due to the large number of chemical precursors and the

transformation of these compounds via subsequent oxidative, photochemical, or reaction pathways. Indeed, a number of methods have been developed to treat the properties of atmospheric organic aerosols, including gas-particle partitioning and chemical aging, which do not deal with individual chemical compounds but account for more general changes in observables such as polarity, volatility, carbon number, and O/C elemental ratio, in order to deal with this complexity.<sup>2–4,9,22–25</sup> Continued improvements in both measurement and modeling should lead to improvements in the accuracy and descriptions of the RIs of organic aerosol. However, a predictive framework that quantifies the optical properties of organic aerosol based on source/composition and changes due to in situ chemical processing would be beneficial.

Given the chemical complexity of organic aerosol, it is challenging to make predictions of fundamental physicochemical properties. This has led to the development of quantitative structure–property relationships (QSPR). This approach is most widely applied to predict vapor pressures in aerosol/gas-phase partitioning models but has been also used to a limited extent to predict the RIs of organic components germane to atmospheric aerosols.<sup>6,26–31</sup> For example, Redmond and Thompson<sup>32</sup> developed and evaluated a QSPR model for the

Received: April 5, 2017

Revised: July 25, 2017

Accepted: July 28, 2017

Published: July 28, 2017

prediction of midvisible organic aerosol RI from molecular formula. They also assessed the applicability of the predictions for organic aerosols derived from chamber experiments and in models of organic aerosol aging. They assumed a linear relationship between the RI of an organic component and the degree of unsaturation ( $\mu$ ), molecular polarizability ( $\alpha$ ), and ratio of the density to molar mass of the component,  $\rho_m/M$ :

$$\text{RI} = 0.031717\mu + 0.0006087\alpha - 3.0227\left(\frac{\rho_m}{M}\right) + 1.38709 \quad (1)$$

There is no physical basis for the assumption of a linear correlation between RI and  $\mu$ ,  $\alpha$ , and  $\rho_m/M$ ; indeed, constraining the relationship of RI to these quantities in this way could lead to systematic inaccuracies in RI predictions. The molar refraction mixing rule has been used alongside bulk aqueous measurements and parametrizations for aqueous solution density to predict the refractive index beyond the solubility limit.<sup>33</sup> However, to the best of our knowledge, a robust and reliable approach for the prediction of RIs of organic components is not currently available.

A plausible approach to represent the large number of different organic compounds present in ambient aerosol with a manageable degree of model complexity is to consider the contributions of organic functional groups instead of individual compounds. Instead of judging a compound as the mixture of different elements, the compound is represented as a mixture of different functional groups. Further, molecular anisotropy can lead to significant intermolecular interactions (H bonding, van der Waals forces, etc.) between components in complex internal aerosol mixtures, and this can be factored in by introducing parameters to represent the interactions between different groups.<sup>31,34–36</sup> The group contribution (GC) approach has been widely introduced into thermodynamic predictions, such as with the well-known UNIFAC model.<sup>37–40</sup> UNIFAC assumes that dividing a molecule into appropriate functional groups, which is then modeled as a “solution of functional groups”, can be used to determine the physicochemical properties of organic molecules. This approach for representing a molecule has been shown to provide an adequate level of accuracy in many instances to predict the real behavior of aqueous–organic and organic mixtures.<sup>37</sup>

In this publication, we develop and evaluate a GC-based method to predict the real part of the RIs of organic components germane to atmospheric aerosols; we do not consider the imaginary part here. Unlike previous models that provided parametrizations of RI based on simple metrics measurable for ambient aerosol,<sup>32,41</sup> this model requires molecular structures to provide predictions of organic compound RIs. It is our purpose in this paper to provide an accurate framework for estimating the RIs of pure organic components using approaches similar to those used in thermodynamic models of, for example, hygroscopic growth. Reduced complexity models based on physically justifiable and validated simplifications will be the subject of future work. Already, we consider that the method may be useful to estimate/predict the relative humidity (RH) and compositional dependence of the RIs of organic aerosols in well-defined laboratory experiments and in models of organic aerosol aging using well-established RI mixing rules.<sup>33,42</sup> Following from a consideration of the deficiencies in simulations and models in previous studies, the model described here considers the role of molecular anisotropy and the presence of significant

intermolecular interactions that might exist in complex internal aerosol mixtures. We will show that the method can provide accurate and reliable pure component organic RIs for further investigation of secondary organic aerosol growth and evolution.

## II. METHOD

The Lorentz–Lorenz relation (also known as the Clausius–Mossotti relation)<sup>43</sup> relates the RI of a substance to the molar polarizability,  $\alpha$ , and density,  $\rho_m$ , of the compound:

$$\alpha = \frac{3}{4\pi} \frac{M}{\rho_m} \left( \frac{n^2 - 1}{n^2 + 2} \right) \quad (2)$$

$\rho_m$  is a mass density of compound with units  $\text{g cm}^{-3}$ ,  $M$  represents molar mass of the compound in  $\text{g mol}^{-1}$ , and  $n$  is the real part of the RI. We first introduce the molar volume,  $V_m$ , into this equation.  $V_m$  can be expressed as

$$V_m = \frac{M}{\rho_m} \quad (3)$$

Therefore, the molar polarizability can be written as

$$\frac{\alpha}{V_m} = \frac{3}{4\pi} \left( \frac{n^2 - 1}{n^2 + 2} \right) \quad (4)$$

The molar polarizability can be expressed in terms of the contributions of different functional groups forming the compound and a stoichiometric weighting for each.<sup>41</sup>

$$\frac{\alpha}{V_m} = \sum_i x_i \frac{\alpha_i}{V_{m,i}} \quad (5)$$

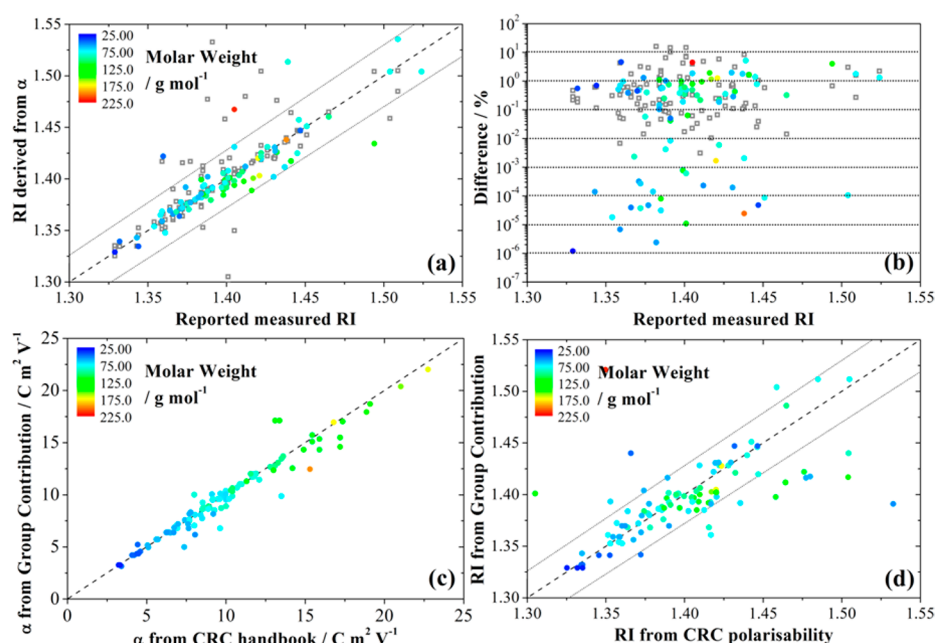
$x_i$  and  $\alpha_i/V_{m,i}$  are the mole fraction and  $\alpha/V_m$  for each functional group  $i$ . Here, we provide a parametrization that leads to the calculation of  $\alpha/V_m$  from this GC approach.

To constrain the contributions of individual functional groups for a GC model, literature values of RIs for a wide selection of compounds are used to provide a training set. Then, values of  $\alpha_i/V_{m,i}$  for each functional group are floated as fit parameters using the relationship

$$\sum_i x_i \frac{\alpha_i}{V_{m,i}} = \frac{3}{4\pi} \left( \frac{n^2 - 1}{n^2 + 2} \right) \quad (6)$$

In this study, functional groups are designated as either main groups or subgroups. Main groups provide a coarse designation of the key functional groups present in the molecule but do not recognize the number of hydrogen atoms bound to carbon atoms, e.g.,  $\text{CH}_n$  (alkyl main group) and  $\text{CH}_n\text{CO}$  (ketone main group). Instead, subgroups resolve the structure of the molecule still further by recognizing the number of H atoms within the main group. For example,  $\text{CH}_n$  is subdivided into  $\text{CH}_3$ ,  $\text{CH}_2$ ,  $\text{CH}$ , and C alkyl chains, and  $\text{CH}_n\text{CO}$  is subdivided into  $\text{CH}_3\text{CO}$  and  $\text{CH}_2\text{CO}$ . This is the same classification of functional groups used in the Aerosol Inorganic–Organic Mixtures Functional groups Activity Coefficients (AIOMFAC-web) equilibrium composition model.<sup>30</sup>

Similar to the implementation in AIOMFAC-web, organic compounds are assigned subgroups representative of the functional groups contained in the organic molecule, with main groups identified automatically.<sup>30</sup> With the GC approach, there are a limited number of main and subgroups that can be used, in principle, to represent an unlimited number of organic



**Figure 1.** (a) Comparison between RIs estimated directly from polarizability in CRC handbook (gray squares), polarizability derived by GC fitting (colored circles mapped by molar weights of corresponding compounds), and the reported value of the measured RI. (b) Percentage differences between estimated RIs from polarizability in CRC handbook and measurements, and between RI from polarizability derived by GC fitting and measurements (symbols and colors same as (a)). (c) Comparison between polarizability from CRC handbook and from GC fitting (colors same as (a)). (d) Comparison between RIs estimated directly from polarizability in CRC handbook and from GC fitting (colors same as (a)).

molecules. Thus, the GC method provides a practical, yet scientifically robust approach to estimate properties for comparison with existing measurements and the predictive capability for molecules for which measurements are not available. This has already been implemented in other studies in models to predict activity coefficients and pure component vapor pressures of atmospherically relevant organic compounds; here, the GC approach is used to determine pure component RIs.<sup>37–40</sup>

In the fitting process, we preset the initial values of  $\alpha_i/V_{m,i}$  of different functional groups as random values. From these preset  $\alpha_i/V_{m,i}$ , we calculate the RIs for compounds in the training set and compare with corresponding RIs reported in the literature. Then we adjust  $\alpha_i/V_{m,i}$  to optimize the predicted RIs and minimize the difference to the literature values using the simplex linear programming solving method with bounds on the variables as implemented as an Excel solver add-in. Three separate GC models were derived to treat atmospheric related compounds containing C, H, N, and O, aromatic compounds, and compounds containing halogens separately. These are referred to as models 1–3, respectively, below. For models 1, 2, and 3, 32, 27, and 25 parameters are required to represent the functional groups (main and sub-) comprising 106, 64, and 64 compounds, respectively. Details of parameters corresponding to various functional groups are shown in Table S1. Clearly, new data on a wider range of compounds are desirable to improve the quality and reliability of the model. Here, we provide the basis for this longer term development.

### III. RELIABILITY IN DERIVATION OF PURE-COMPONENT RIS FOR COMPOUNDS IN TRAINING SET

The results of fitting 106 different organic compounds (the training set) for model 1 are shown in Figure 1. Details of compounds in model 1 are presented in Table S2. Functional

groups considered in the fitting are the unsaturated carbon bond, OH, NH<sub>x</sub>, NO<sub>2</sub>, and  $-\text{C}(=\text{O})-$  functional groups. Details of all compounds in our training set including name, molecular structure, polarizability, and RIs from different methods and associated absolute errors are provided in Table S2 in the Supporting Information. All these compounds are potential organic compounds identified as occurring in the atmospheric environment. For comparison, the molar polarizability of each organic compound in this training set is also known from the CRC Handbook of Chemistry and Physics.<sup>44</sup> Pure component RIs are known from previous measurements and the values adopted are those reported by Sigma-Aldrich in their catalogue. Indeed, the existence of both molar polarizability data and RIs are key conditions for the selection of a compound to use in the training set and necessarily limit the available data for the training set. Molar weights of compounds in the training set are between  $\sim 25$  and  $\sim 225$  g mol<sup>−1</sup>.

As observed in Figure 1a, although data points (each corresponding to an individual compound) do exhibit scatter around the 1:1 line, RIs derived from the GC fitting give much better consistency with measured RIs than when derived directly from the molecular polarizabilities reported in the CRC Handbook of Chemistry and Physics and eq 4. The quality of the fitting is also shown as the percentage difference between a RI from the GC fitting or CRC polarizability data and the measurement in Figure 1b. The average differences and the standard deviation of RIs are illustrated in Table 1; the percentage differences reported in Figure 1 are calculated from the difference between the modeled RI and the value reported in the literature, relative to the reported value. A direct comparison between polarizabilities and RIs from GC fitting and CRC polarizabilities is provided in Figures 1c,d.

In an effort to further characterize the performance of the model, we have considered deviations from the measured RIs with compound functionalities. In Figure S1, we summarize the



**Table 1. Error and Standard Deviation of Predictions of the RIs of Compounds in the Training Set (Models 1–3) When Compared with Measured RIs<sup>a</sup>**

models	from GC method		from polarizability in <i>The CRC Handbook</i>	
	error (%)	standard derivation (%)	error (%)	standard derivation (%)
model for atmospheric related compounds (model 1)	0.58	0.94	1.51	1.38
model for aromatic compounds (model 2)	0.36	1.09	1.14	1.43
model for compounds containing halogen (model 3)	0.30	0.72	1.68	0.86

<sup>a</sup>The predictions are derived from the GC model and from polarizabilities reported in the *CRC Handbook of Chemistry and Physics*.

influence of different functional groups on pure component RIs and the accuracy with which the GC model represents the literature data. The figure is divided into functional group categories of alkenes/alkynes, alcohols, amides/amines, nitro compounds, and aldehydes/ketones/carboxylic acids/esters. There are 22 compounds containing C–C double or triple bonds (alkenes, alkyne), 33 compounds containing hydroxyl (alcohols), 18 compounds containing an amino group (amide/amines), six compounds containing nitril groups (nitro compound), and 54 compounds containing a carbonyl group (aldehydes/ketones/carboxylic acids/ester). For each compound class, an average absolute error has been calculated to consider whether systematic trends exist between certain compound classes. The average differences and the standard deviation of RIs containing these different groups are reported in Table 2.

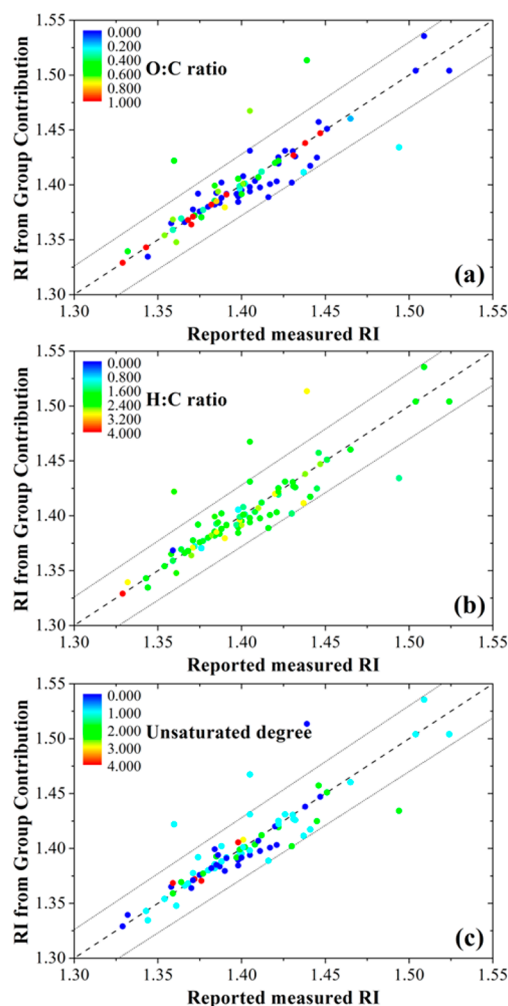
**Table 2. Error and Standard Derivation of RIs Calculated from GC Model When Compared to Measured Values for Compounds in the Training Set for Model 1**

different functional groups	error (%)	standard derivation (%)
C=C or C≡C	0.41	1.21
OH	0.18	0.96
NH <sub>x</sub>	0.57	0.50
NO <sub>2</sub>	0.04	0.47
C=O	0.19	0.98

In order to evaluate the possible influence of the oxidative aging of aerosol on optical properties, we compare the GC model predictions with measurements highlighting the dependence on O/C ratio, H/C ratio, and degree of unsaturation in Figure 2a, b, and c, respectively (all data are the same as Figure 1). The limited data used in this study do not reveal systematic trends in RI with O/C and H/C ratios. Competing factors, such as component density, MW, multiple functional group types, molecular structure, etc., which each play a role in contributing to the value of RI, disguise any general trends across the coarse range of molecular compounds examined.

It has also been noted in a previous approach<sup>32</sup> that there is a linear relationship between RIs and degree of unsaturation, defined as follows

$$\mu = (\#C + 1) - \frac{\#H - \#N}{2} \quad (7)$$

**Figure 2.** Consistency between reported and predicted RIs of atmospheric related compounds used in the training set for model 1 classified according to (a) O/C ratio, (b) H/C ratio, and (c) degree of unsaturation.

in which  $\mu$  is the degree of unsaturation, while  $\#C$ ,  $\#H$ , and  $\#N$  are simply the number of atoms of each element in the molecular formula.<sup>32</sup> Similar to the potential general trends in RI with O/C and H/C considered above, it is clear in Figure 2c that there is little correlation between the RI and the degree of unsaturation for the limited number of compounds considered here. As discussed above, real RIs can be influenced by the complex interplay of various factors such as component density, MW, and functional group type and molecular structure. Any general trends previously observed may not be observable if all of these vary from compound to compound and if the coverage of compounds (size of the training set) is too small.<sup>45–47</sup> From Figure 2, the model shows good accuracy and reliability independent of degree of oxidation and aging.

We have also evaluated the performance of the GC method for the training sets of 64 aromatic organic compounds (model 2) and 64 organic compounds containing halogen atoms (model 3). Molar weights of aromatic organic compounds in the training set are between  $\sim 70$  g mol<sup>−1</sup> and  $\sim 270$  g mol<sup>−1</sup>; similarly, molar weights of organic compounds containing halogens in the training set are between  $\sim 70$  and  $\sim 270$  g mol<sup>−1</sup>. Detailed information for these two models is provided in the Supporting Information; here, it is sufficient to state that

similar levels of accuracy and similar trends are observed to the data used in model 1.

#### IV. PREDICTIONS OF PURE COMPONENT RIS FOR A BENCHMARK OXIDATION MECHANISM

In order to evaluate the performance of the GC method, we report estimations of the RIs of the ozonolysis products of maleic acid (MA), a benchmark system chosen based on our previous work.<sup>48,49</sup> We have chosen this system to benchmark the GC model here, having previously reported the use of the models provided by Cappa et al.<sup>41</sup> and Redmond and Thompson<sup>32</sup> to rationalize the measured RIs of oxidized MA aerosol determined from aerosol optical tweezers measurements. The details of the MA ozonolysis products considered in our previous study are shown in Table S5, along with the estimates of the pure component RIs.<sup>48</sup> The first method, proposed by Cappa et al.,<sup>41</sup> is based on the RIs of heterogeneous oxidative aging products of squalane (a C<sub>30</sub>-saturated hydrocarbon) and azelaic acid (a C<sub>9</sub> dicarboxylic acid) by the hydroxyl radical (OH) and is based on a parametrization for the RIs derived from the elemental composition of dry aerosol. This allows estimation of the RIs of organic aerosol without knowledge of the specific molecular identity of the aerosol components. A second method, proposed by Redmond and Thompson,<sup>32</sup> is based on a quantitative structure–property relationship. It allows the prediction of the RIs of components of secondary organic aerosol from their molecular formula and density of 111 compounds to provide a parametrization for estimation of the RI of a compound.

In Figure 3, we report estimates of the RIs of the products from MA ozonolysis using these two previously reported methods and the GC method.<sup>48</sup> Details of these compounds are presented in Table S5. Uncertainties of RIs estimated using the methods of Cappa et al.<sup>41</sup> and Redmond and Thompson<sup>32</sup>

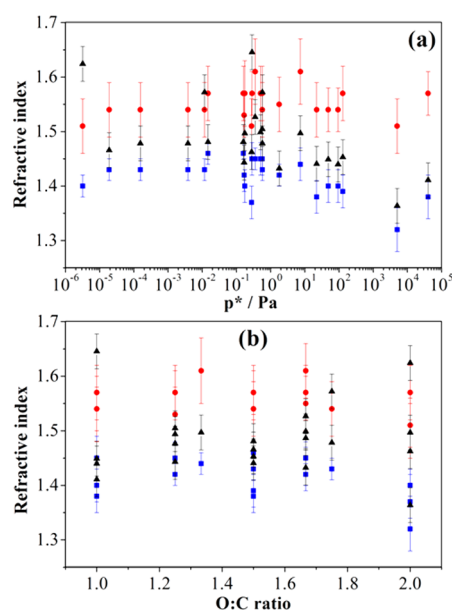
come from the corresponding studies, while uncertainties of RIs estimated from group contribution are derived from the standard deviations for the different GC parameters. The predicted values of the RIs estimated by using the approach of Cappa and co-workers are consistently higher than those estimated by the method of Redmond and Thompson. Consistency between the treatments is quite poor with predicted values for the same compound spanning an implausibly large range of as much as 1.35 to 1.6. Despite this, some similar relative trends between compounds do exist between models. Further, predictions from the GC method are mostly located between predictions from the Cappa et al.<sup>41</sup> and Redmond and Thompson models.<sup>32</sup> The calculated RIs from both methods can be compared to the estimate of  $1.481 \pm 0.001$  for the RI of a subcooled melt of pure MA determined in optical tweezer experiments.<sup>48</sup> The RIs of MA estimated using the methods of Cappa et al.<sup>41</sup> and Redmond and Thompson<sup>32</sup> are  $1.57 \pm 0.04$  and  $1.45 \pm 0.03$ , respectively, while the RI estimated using the GC method is  $1.479 \pm 0.03$ , very close to the optical tweezers measurement of a droplet that exists as a subcooled liquid. MA is not in the original training set for model 1. The crystalline RI of MA is 1.509.<sup>49</sup> Predictions from the two earlier methods are based on parametrizations using compounds much less oxygenated than MA and the potential ozonolysis products; by contrast, the GC method is parametrized using compounds with a wide range of O:C ratio (see Figure 2 as an example). On the basis of our earlier measurements,<sup>48</sup> the RI of oxidized MA aerosol under dry conditions can be expected to fall in a range above 1.48, consistent with the predictions of the model presented here. For the work of Cappa et al.,<sup>41</sup> the maximum O/C ratio used in the fitting was 0.8; potential MA aerosol ozonolysis products could have O/C ratios between 1 and 2.

In order to compare results from the GC fitting introduced in this study and results from previous simulations, we compare RIs of all 234 compounds in our training sets (all results illustrated in Figure 1 and Figures S2 and S3) together with the RIs of all 111 compounds used in the training set by Redmond and Thompson<sup>32</sup> in Figure 4. Details of compounds in Redmond and Thompson's work<sup>32</sup> are presented in Table S6. Predicted RIs from the GC method provide a much more accurate representation of the RIs of all 335 compounds, including the compounds used by Redmond and Thompson to form their training set. This is apparent from the percentage differences between RIs from GC or Redmond and Thompson method and RIs from measurements as shown in Figure 4b.

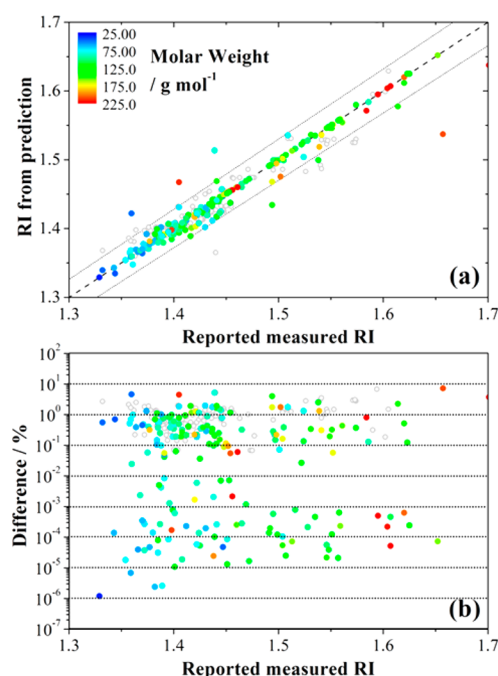
A fuller evaluation of the performance of our GC method in reproducing the component RIs in the training set used by Redmond and Thompson, and the performance of the Redmond and Thompson model in replicating our training set is provided in the Supporting Information. The average differences and the standard deviations for comparisons of all RIs using training sets in different combinations are reported in Table 3.

#### V. COMPARISON BETWEEN GC PREDICTIONS AND PURE COMPONENT RI PREDICTIONS FROM APPLICATION OF MOLAR REFRACTION MIXING RULE TO BULK-PHASE MEASUREMENTS OF AQUEOUS SOLUTIONS

The results of predictions for 38 different organic compounds and their comparison with estimations from the application of



**Figure 3.** Estimated RIs of ozonolysis products of maleic acid aerosol under humidified conditions. RIs from the methods of Redmond and Thompson<sup>32</sup> (blue squares) and Cappa et al.<sup>41</sup> (red circles) and derived from the GC method (black triangles). The estimated RIs are shown against (a) the estimated vapor pressure of the component and (b) the O/C ratio.

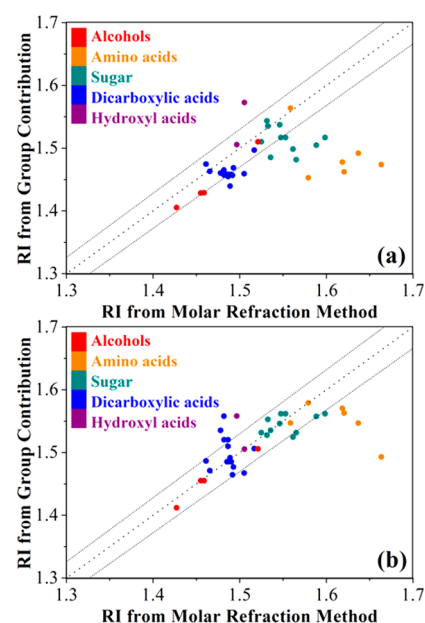


**Figure 4.** Comparison of the accuracy of predictions of RI of training set compounds from Redmond and Thompson's work<sup>32</sup> and RI of training set compounds from this study. (a) Comparison of RI from prediction and measurements in Redmond and Thompson's work<sup>32</sup> (gray hollow squares) and comparison of RI from GC model and reported measurements (colored solid circles mapped by organics molar weights). (b) Differences between RI from prediction and measurements in Redmond and Thompson's work<sup>32</sup> (gray hollow squares) and differences of RI from GC model and reported measurements (colored solid circles mapped by organics molar weights).

**Table 3. Error and Standard Derivation of RIs for Training Set Compounds from Redmond and Thompson's Work and the RIs of Training Set Compounds from This Study**

fit	from GC method		from Redmond and Thompson's work	
	error (%)	standard derivation (%)	error (%)	standard derivation (%)
fit for compounds in two training sets combined	0.44	0.74	1.07	1.03
fit for compounds used in training set of Redmond and Thompson's study	1.88	1.07	1.07	1.03
fit for compounds in training set in this study	0.44	0.74	2.94	1.59

molar refraction mixing rule to bulk aqueous solution measurements to estimate pure component values are shown in Figure 5. These measurements have been previously reported and discussed by us, and we refer the reader to our previous publications for more information and Table S7.<sup>33,50</sup> All of these compounds are potential components of ambient aerosol. These substances include the following chemical functionalities: alcohols, amino acids, sugars, dicarboxylic acids, and hydroxyl acids. Molar polarizabilities are not available for any of these organic compounds, and so predictions can only be made using the GC approach. As observed in Figure 5a, RIs derived from the GC fitting do not yield equally accurate RIs across all compound classes. While predicted RIs of the



**Figure 5.** (a) Comparison between RIs of organic compounds from application of molar refraction mixing rule to bulk phase measurements and predictions from the GC method with color identifying the organic functionality of the molecule. (b) Same as (a) but following a refinement of the GC parameters using the molar refraction method derived data in addition to the original training set used to fit model 1.

alcohols and dicarboxylic acids are approximately consistent with measured RIs, those for amino acids, sugars, and hydroxyl acids can be very different from measured RIs, especially amino acids. The average differences and the standard deviation of RI containing different groups are summarized in Table 4.

**Table 4. Error and Standard Derivation of RIs of Organic Compounds Determined from Fitting Bulk Aqueous Solution Phase Data to Molar Refraction<sup>33</sup>**

selected model	compounds	error (%)	standard derivation (%)
model for nonaromatic compounds	alcohols	0.58	0.54
	amino acids	8.60	8.30
	sugars	3.00	2.50
	dicarboxylic acids	2.10	1.90
	hydroxyl acids	2.60	3.70
selected model	compounds	error (%)	standard derivation (%)
model for nonaromatic compounds refined with additional compounds using molar refraction	alcohols	0.60	0.50
	amino acids	3.80	3.70
	sugars	1.10	0.90
	dicarboxylic acids	1.60	1.50
	Hydroxyl acids	2.10%	2.90%

In order to provide a refined GC model using the compound functionalities from our molar refraction measurements, we have included these 38 compounds with the 106 organic compounds previously used in the training set for model 1. The results of this refined model are presented in Figure 5b, illustrating the improvement in the predicted values. The average differences and the standard deviations of RIs



containing different groups are also illustrated in Table 4. This is a first example of the model retraining that can be pursued to improve the accuracy and generality of the model as more data become available.

In summary, we report a group contribution model for predicting the RIs of pure organic components relevant for rationalizing the optical properties of atmospheric organic aerosol. We suggest that this model will be invaluable in interpreting laboratory based measurements of aerosol processes. Future developments of the model will focus on developing the simplicity of model, removing the need for structural information for the compounds considered while retaining robust and verifiable estimates of optical properties.

## ■ ASSOCIATED CONTENT

### Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.est.7b01756.

Refractive indices for all compounds used in training sets, group contribution coefficients for all models, and figures reporting errors in model predictions (PDF)

## ■ AUTHOR INFORMATION

### Corresponding Author

\*E-mail: j.p.reid@bristol.ac.uk.

### ORCID

Yun-hong Zhang: 0000-0002-2529-3538

Jonathan P. Reid: 0000-0001-6022-1778

### Notes

The authors declare no competing financial interest.

All data used and reported in this publication are provided in the Supporting Information.

## ■ ACKNOWLEDGMENTS

J.P.R. acknowledges the Natural Environment Research Council for support through Grant No. NE/L006901/1. Y.-h.Z. acknowledges the National Natural Science Foundation of China (Nos. 21473009 and 91544223). C.C. acknowledges support from the China Scholarship Council for 2012 International Joint PhD Project. A.M. acknowledges the EPSRC for DTA.

## ■ REFERENCES

- (1) IPCC. *IPOC Climate Change 2007-The Physical Science Basis: Working Group Contribution to the Fourth Assessment Report of the IPCC*; Cambridge University Press: Cambridge, UK, 2007.
- (2) Jimenez, J. L.; Canagaratna, M. R.; Donahue, N. M.; Prevot, A. S.; Zhang, Q.; Kroll, J. H.; Decarlo, P. F.; Allan, J. D.; Coe, H.; Ng, N. L. Evolution of Organic Aerosols in the Atmosphere. *Science* **2009**, 326, 1525–9.
- (3) Donahue, N. M.; Robinson, A. L.; Stanier, C. O.; Pandis, S. N. Coupled Partitioning, Dilution, and Chemical Aging of Semivolatile Organics. *Environ. Sci. Technol.* **2006**, 40, 2635–43.
- (4) Donahue, N. M.; Robinson, A. L.; Pandis, S. N. Atmospheric Organic Particulate Matter: From Smoke to Secondary Organic Aerosol. *Atmos. Environ.* **2009**, 43, 94–106.
- (5) Ng, N. L.; Canagaratna, M. R.; Zhang, Q.; Jimenez, J. L.; Tian, J.; Ulbrich, I. M.; Kroll, J. H.; Docherty, K. S.; Chhabra, P. S.; Bahreini, R.; et al. Organic Aerosol Components Observed in Northern Hemispheric Datasets From Aerosol Mass Spectrometry. *Atmos. Chem. Phys.* **2010**, 10, 4625–4641.
- (6) Pankow, J. F.; Asher, W. E. SIMPOL.1: A Simple Group Contribution Method for Predicting Vapor Pressures and Enthalpies of Vaporization of Multifunctional Organic Compounds. *Atmos. Chem. Phys.* **2008**, 8, 2773–2796.
- (7) George, I. J.; Abbatt, J. P. D. Chemical Evolution of Secondary Organic Aerosol From OH-initiated Heterogeneous Oxidation. *Atmos. Chem. Phys. Discuss.* **2010**, 10, 3265–3300.
- (8) Charlson, R. J.; Schwartz, S. E.; Hales, J. M.; Cess, R. D.; Coakley, J. A.; Hansen, J. E.; Hofmann, D. J. Climate Forcing by Anthropogenic Aerosols. *Science* **1992**, 255, 423–430.
- (9) Pankow, J. F.; Barsanti, K. C. The Carbon Number-Polarity Grid: A Means to Manage the Complexity of the Mix of Organic Compounds When Modelling Atmospheric Organic Particulate Matter. *Atmos. Environ.* **2009**, 43, 2829–2835.
- (10) Kanakidou, M.; Seinfeld, J. H.; Pandis, S. N.; Barnes, I. Organic Aerosol and Global Climate Modelling: A Review. *Atmos. Chem. Phys.* **2005**, 5, 1053–1123.
- (11) Dickerson, R. R.; Kondragunta, S.; Stenchikov, G.; Civerolo, K. L.; Doddridge, B. G.; Holben, B. N. The Impact of Aerosols On Solar Ultraviolet Radiation and Photochemical Smog. *Science* **1997**, 278, 827–30.
- (12) Jacobson, M. Z. Isolating Nitrated and Aromatic Aerosols and Nitrated Aromatic Gases as Sources of Ultraviolet Light Absorption. *J. Geophys. Res. Atmos.* **1999**, 104, 3527–3542.
- (13) Murphy, D. M.; Czicz, D. J.; Froyd, K. D.; Hudson, P. K.; Matthew, B. M.; Middlebrook, A. M.; Peltier, R. E.; Sullivan, A.; Thomson, D. S.; Weber, R. J. Single-Particle Mass Spectrometry of Tropospheric Aerosol Particles. *J. Geophys. Res.* **2006**, 111, 869–892.
- (14) Zhang, Q.; Jimenez, J. L.; Canagaratna, M. R.; Allan, J. D.; Coe, H.; Ulbrich, I.; Alfarra, M. R.; Takami, A.; Middlebrook, A. M.; Sun, Y. L.; et al. Ubiquity and Dominance of Oxygenated Species in Organic Aerosols in Anthropogenically-Influenced Northern Hemisphere Midlatitudes. *Geophys. Res. Lett.* **2007**, 34, L13801.
- (15) Abo Rizeq, A.; Erlick, C.; Dinar, E.; Rudich, Y. Optical Properties of Absorbing and Non-Absorbing Aerosols Retrieved by Cavity Ring Down (CRD) Spectroscopy. *Atmos. Chem. Phys.* **2007**, 7, 1523–1536.
- (16) Schnaiter, M.; Horvath, H.; Möhler, O.; Naumann, K. H.; Saathoff, H.; Schöck, O. W. UV-VIS-NIR Spectral Optical Properties of Soot and Soot-Containing Aerosols. *J. Aerosol Sci.* **2003**, 34, 1421–1444.
- (17) Barkey, B.; Paulson, S. E.; Chung, A. Genetic Algorithm Inversion of Dual Polarization Polar Nephelometer Data to Determine Aerosol Refractive Index. *Aerosol Sci. Technol.* **2007**, 41, 751–760.
- (18) Yu, Y.; Ezell, M. J.; Zelenyuk, A.; Imre, D.; Alexander, L.; Ortega, J.; D'Anna, B.; Harmon, C. W.; Johnson, S. N.; Finlayson-Pitts, B. J. Photooxidation of  $\alpha$ -Pinene at High Relative Humidity in the Presence of Increasing Concentrations of  $\text{NO}_x$ . *Atmos. Environ.* **2008**, 42, 5044–5060.
- (19) Kim, H.; Barkey, B.; Paulson, S. Real Refractive Indices of  $\alpha$ - and  $\beta$ -Pinene and Toluene Secondary Organic Aerosols Generated From Ozonolysis and Photo-Oxidation. *J. Geophys. Res. Atmos.* **2010**, 115, D24212.
- (20) Langyona, N.; Rudich, Y.; Mentel, T. F.; Buchholz, A. The Chemical and Microphysical Properties of Secondary Organic Aerosols from Holm Oak Emissions. *Atmos. Chem. Phys.* **2010**, 10, 7253–7265.
- (21) Nakayama, T.; Matsumi, Y.; Sato, K.; Imamura, T.; Yamazaki, A.; Uchiyama, A. Laboratory Studies On Optical Properties of Secondary Organic Aerosols Generated During the Photooxidation of Toluene and the Ozonolysis of  $\alpha$ -Pinene. *J. Geophys. Res. Atmos.* **2010**, 115, 9–12.
- (22) Heald, C. L.; Kroll, J. H.; Jimenez, J. L.; Docherty, K. S.; Decarlo, P. F.; Aiken, A. C.; Chen, Q.; Martin, S. T.; Farmer, D. K.; Artaxo, P. A simplified description of the evolution of organic aerosol composition in the atmosphere. *Geophys. Res. Lett.* **2010**, 37, L08803.
- (23) Presto, A. A.; Gordon, T. D.; Robinson, A. L. Primary to secondary organic aerosol: evolution of organic emissions from mobile combustion sources. *Atmos. Chem. Phys.* **2014**, 14, 5015–5036.
- (24) Fooshee, D. R.; Nguyen, T. B.; Nizkorodov, S. A.; Laskin, J.; Laskin, A.; Baldi, P. COBRA: a computational brewing application for



predicting the molecular composition of organic aerosols. *Environ. Sci. Technol.* **2012**, *46*, 6048–6055.

(25) Wei, Y.-Y.; Cao, T. T.; Thompson, J. E. The chemical evolution physical properties of organic aerosol: A molecular structure based approach. *Atmos. Environ.* **2012**, *62*, 199–207.

(26) Asher, W. E.; Pankow, J. F. Vapor Pressure Prediction for Alkenoic and Aromatic Organic Compounds by a UNIFAC-based Group Contribution Method. *Atmos. Environ.* **2006**, *40*, 3588–3600.

(27) Clegg, S. L.; Kleeman, M. J.; Griffin, R. J.; Seinfeld, J. H. Effects of Uncertainties in the Thermodynamic Properties of Aerosol Components in an Air Quality Model - Part 2: Predictions of the Vapour Pressures of Organic Compounds. *Atmos. Chem. Phys.* **2008**, *8*, 1087–1103.

(28) Barley, M.; Topping, D. O.; McFiggans, G.; Jenkin, M. E. Sensitivities of the Absorptive Partitioning Model of Secondary Organic Aerosol Formation to the Inclusion of Water. *Atmos. Chem. Phys.* **2009**, *9*, 2919.

(29) Girolami, G. S. A Simple Back of the Envelope Method for Estimating the Densities and Molecular Volumes of Liquids and Solids. *J. Chem. Educ.* **1994**, *71*, 962–964.

(30) Zuend, A.; Marcolli, C.; Booth, A. M.; Lienhard, D. M.; Soonsin, V.; Krieger, U. K.; Topping, D. O.; McFiggans, G.; Peter, T.; Seinfeld, J. H. New and Extended Parameterization of the Thermodynamic Model AIOMFAC: Calculation of Activity Coefficients for Organic-Inorganic Mixtures Containing Carboxyl, Hydroxyl, Carbonyl, Ether, Ester, Alkenyl, Alkyl, and Aromatic Functional Groups. *Atmos. Chem. Phys.* **2011**, *11*, 9155–9206.

(31) Decesari, S.; Facchini, M. C.; Matta, E.; Mircea, M.; Fuzzi, S.; Chughtai, A. R.; Smith, D. M. Water Soluble Organic Compounds Formed by Oxidation of Soot. *Atmos. Environ.* **2002**, *36*, 1827–1832.

(32) Redmond, H.; Thompson, J. E. Evaluation of a Quantitative Structure-Property Relationship (QSPR) for Predicting Mid-Visible Refractive Index of Secondary Organic Aerosol (SOA). *Phys. Chem. Chem. Phys.* **2011**, *13*, 6872–6882.

(33) Cai, C.; Miles, R. E. H.; Cotterell, M. I.; Marsh, A.; Rovelli, G.; Rickards, A. M. J.; Zhang, Y.-H.; Reid, J. P. Comparison of Methods for Predicting the Compositional Dependence of the Density and Refractive Index of Organic–Aqueous Aerosols. *J. Phys. Chem. A* **2016**, *120*, 6604–6617.

(34) Shapiro, E. L.; Szprengiel, J.; Sareen, N.; Jen, C. N.; Giordano, M. R.; McNeill, V. F. Light-Absorbing Secondary Organic Material Formed by Glyoxal in Aqueous Aerosol Mimics. *Atmos. Chem. Phys.* **2009**, *9*, 2289–2300.

(35) Chang, J. L.; Thompson, J. E. Characterization of Colored Products Formed During Irradiation of Aqueous Solutions Containing H<sub>2</sub>O<sub>2</sub> and Phenolic Compounds. *Atmos. Environ.* **2010**, *44*, 541–551.

(36) Rincon, A. G.; Guzman, M. I.; Hoffmann, M. R.; Colussi, A. J. Optical Absorptivity versus Molecular Composition of Model Organic Aerosol Matter. *J. Phys. Chem. A* **2009**, *113*, 10512–10520.

(37) Fredenslund, A.; Jones, R. L.; Prausnitz, J. M. Group-Contribution Estimation of Activity Coefficients in Nonideal Liquid Mixtures. *AIChE J.* **1975**, *21*, 1086–1099.

(38) Hansen, H. K.; Rasmussen, P.; Fredenslund, A.; Schiller, M.; Gmehling, J. Vapor-Liquid Equilibria by UNIFAC Group Contribution. 5. Revision and Extension. *Ind. Eng. Chem. Res.* **1991**, *30*, 2352–2355.

(39) Kiepe, J.; Noll, O.; Gmehling, J. Modified LIQUAC and Modified LIFACA Further Development of Electrolyte Models for the Reliable Prediction of Phase Equilibria with Strong Electrolytes. *Ind. Eng. Chem. Res.* **2006**, *45*, 2361–2373.

(40) Yan, W.; Toppoff, M.; Rose, C.; Gmehling, J. Prediction of Vapor–Liquid Equilibria in Mixed-Solvent Electrolyte Systems Using the Group Contribution Concept. *Fluid Phase Equilib.* **1999**, *162*, 97–113.

(41) Cappa, C. D.; Che, D. L.; Kessler, S. H.; Kroll, J. H.; Wilson, K. R. Variations in Organic Aerosol Optical and Hygroscopic Properties upon Heterogeneous OH Oxidation. *J. Geophys. Res.* **2011**, *116*, D15204.

(42) Cai, C.; Stewart, D. J.; Reid, J. P.; Zhang, Y.-H.; Ohm, P.; Dutcher, C. S.; Clegg, S. L. Organic Component Vapor Pressures and Hygroscopicities of Aqueous Aerosol Measured by Optical Tweezers. *J. Phys. Chem. A* **2015**, *119*, 704–718.

(43) Born, M.; Wolf, E. *Principles of Optics*; Pergamon Press, 1980; p 986.

(44) Lide, D. R. *CRC Handbook of Chemistry and Physics*, 44th ed.; Chemical Rubber Company: Boca Raton, FL, 1962.

(45) Eisenlohr, F. *Spektrochemie organischer Verbindungen*. *Z. Phys. Chem.* **1911**, *75*, 585.

(46) Schuhmacher, J. P.; Wibaut, J. P. Synthesis, refractometric constants, and some other physical properties of a homologous series of branched alkenes. *Recl. Trav. Chim. Pays-Bas* **1953**, *72*, 1037–1075.

(47) Geldof, H.; Wibaut, J. P. Refractometric investigations on unbranched alkenes with a terminal double bond. The refraction equivalent of the >C=C< bond. *Recl. Trav. Chim. Pays-Bas* **1948**, *67*, 105–131.

(48) Dennis-Smith, B. J.; Marshall, F. H.; Miles, R. E. H.; Preston, T. C.; Reid, J. P. Volatility and Oxidative Ageing of Aqueous Maleic Acid Aerosol Droplets and the Dependence on Relative Humidity. *J. Phys. Chem. A* **2014**, *118*, 5680–5691.

(49) Marshall, F. H.; Miles, R. E. H.; Song, Y.-C.; Ohm, P. B.; Power, R. M.; Reid, J. P.; Dutcher, C. S. Diffusion and reactivity in ultraviscous aerosol and the correlation with particle viscosity. *Chem. Sci.* **2016**, *7*, 1298–1308.

(50) Marsh, A.; Miles, R. E. H.; Rovelli, G.; Cowling, A. G.; Nandy, L.; Dutcher, C. S.; Reid, J. P. Influence of Organic Compound Functionality on Aerosol Hygroscopicity: Dicarboxylic Acids, Alkyl-Substituents, Sugars and Amino Acids. *Atmos. Chem. Phys.* **2017**, *17* (9), 5583–5599.